

ROTATOR: FLEXIBLE DISTRIBUTION OF DATA ACROSS SENSORY CHANNELS

Juliana Cherston

Responsive Environments Group
MIT Media Lab
75 Amherst Street
Cambridge MA, 02139 USA
cherston@media.mit.edu

Joseph A. Paradiso

Responsive Environments Group
MIT Media Lab
75 Amherst Street
Cambridge MA, 02139 USA
joep@media.mit.edu

ABSTRACT

‘Rotator’ is a web-based multisensory analysis interface that enables users to shift streams of multichannel scientific data between their auditory and visual sensory channels in order to better discern structure and anomaly in the data. This paper provides a technical overview of the Rotator tool as well as a discussion of the motivations for integrating flexible data display into future analysis and monitoring frameworks. An audio-visual presentation mode in which only a single stream is visualized at any given moment is identified as a particularly promising alternative to a purely visual information display mode. Auditory and visual excerpts from the interface are available at <http://resenv.media.mit.edu/rotator>.

1. INTRODUCTION

While walking down the streets of a bustling city, we are unfailingly submerged in a three-dimensional, 360° sonic cacophony of urban life that is commonly referenced in academic literature as an *auditory scene*. Our eyes, cued by the sounds around us, dart purposefully around the scene. For example, when a car screeches, we impulsively jolt our heads to bring the source of the sudden noise into our field-of-view. In this way, we are accustomed to using visual and sonic information in tandem, placing our eyes where they are most beneficial and tuning our attention to precise features embedded within an auditory scene. However, unlike audio, visual input is directional, providing a field-of-view of only 130-135° vertically and 200° horizontally with respect to the center of one’s gaze [1]. It is therefore worth considering whether the brain’s natural methods for distributing data across our auditory and visual sensory channels can aid in the design of a productive high-dimensional data analysis framework.

2. OVERVIEW OF ROTATOR TOOL DESIGN

The Rotator tool is a web-based data analysis framework that enables flexible distribution of scientific data between the user’s eyes and ears. It is designed to study how our perception of the structure of a dataset can be informed and potentially optimized by deliberate manipulation of the presentation modes of its constituent data streams. The expected workflow is as follows: a user loads in a

dataset to be analyzed. The streams within the dataset are assumed to have a geometric interpretation and are displayed on a suitable map. For example, the data may be derived from a physical sensor network or may represent different stages of a process that can be laid out as a schematic. The user begins to explore the data by moving and resizing auditory and visual windows around the node network (see Figure 1 as an example). Nodes falling within one of these windows will be displayed accordingly, either as spatialized audio (where sonified streams are distributed with respect to the center of the sliding audio window), or as a series of line/scatter plots rendered adjacent to the node map. The user may find that listening to a spatialized sonification of all of the data nodes while only looking at visual representations of a few nodes at a time is most comfortable (Figure 1, center). Alternatively, the user may wish to divide the available data streams between his or her two sensory modes (Figure 1, right). When both auditory and visual modes are enabled, a vertical red line slides across each visual plot to provide an approximate indication to users of the data region currently being sonified (Figure 5, left column).

In order to further explore the data, the user can adjust the data-to-audio mapping settings in the audio control panel (Figure 2), shift the range of data within each stream that is currently visualized or sonified, view the FFT of any individual data stream, enable a ‘play’ mode that simulates a real-time monitoring scenario, or cluster the data nodes, among other available controls.

3. THEORY AND PRIOR ART

3.1. Multichannel Monitoring and Expectancy Violation

Rotator relies on the human brain’s capacity for auditory stream segregation and multi-channel visual monitoring. However, studies have repeatedly shown that humans can only monitor four to five items simultaneously before experiencing cognitive overload [2, 3]. Generally, research addressing this perceptual limitation centers on methods for pre-processing, filtering, and auto-labeling data in order to render only the most critical information to the user. That said, many such commonplace approaches, e.g. dimensionality reduction algorithms, inherently obscure the meaning of each individual data stream. It is therefore useful to recognize that the visual system and the auditory system can serve complementary roles in data display. For example, while visual displays work well for providing detailed renderings of local areas of a dataset, aural scanning is better suited for finding regions of interest with short temporal durations in a large dataset. In each case, the user is particularly sensitive to anomaly. For example, in audio, Cariani



This work is licensed under Creative Commons Attribution Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0>

<https://doi.org/10.21785/icad2017.009>

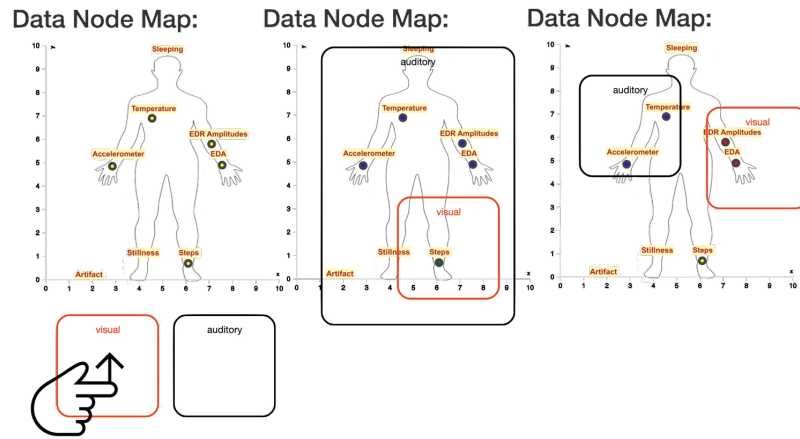


Figure 1: Three sample configurations for Rotator's 'Data Node Map' interface applied in this case to data from a wearable biosensing system. Users resize and slide auditory and visual windows around a map of data stream nodes in order to dictate in which sensory channel each node will be displayed

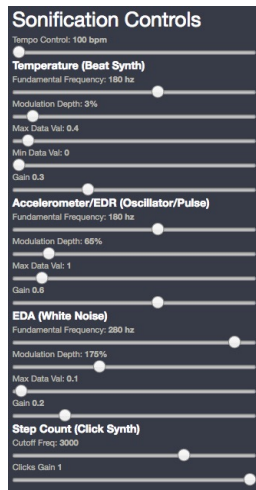


Figure 2: Sample configuration for Rotator's left-hand sonification control panel pictured for the wearable biosensing system application area. The user can adjust the overall tempo as well as data-to-audio mapping parameters for each synthesizer. For example, in the case of the beat effect synthesizer used to sonify body temperature, the user can modify fundamental frequency, modulation depth, min/max data values expected in the stream, and gain

and Micheyl write that the human benefits from expectancy violation effects in the auditory cortex: 'On all timescales, repeating patterns of sound and their evoked auditory events build up strong representational expectancies of their continuation. This effect is created even with arbitrary and highly artificial repeating sound patterns' [4]. In Rotator, users can probe expectancy-violating auditory effects in the visual domain and vice-versa. Expectancy violating auditory events may include the fleeting increase in frequency of a beat effect synthesizer, or the pulsation of an envelope synthesizer, to name some examples.

3.2. Auditory Scene Analysis and Stream Segregation

Auditory Scene Analysis (ASA) describes a set of heuristics that model the organization of incoming auditory data into a scene around the listener. Within an auditory scene of discrete and sufficiently segregated sound sources, the listener's attention can shift from source to source [5]. Rotator draws from prior work in stream segregation, notably in the heavy use of spatialization, pitch segregation, and the overlay of continuous and intermittent sounds.

Pioneering work on auditory scene analysis was conducted by Bregman in 1994 [6]. Cariani's biological framing of 'expectancy violation' in the previous section is qualitatively similar to Bregman's 'old-plus-new' heuristic: when new components are abruptly added to a spectrum of partials, the ASA system is skilled at deducing which partials are a continuation of the previous signal and which are newly added. The newly added partials are perceived as a separate sound [6].

Efforts have been made to model perceptual segregation of streams in terms of common auditory parameters. For example, the van Noorden Diagram maps the perception of two tones with respect to their pitch difference and intertone onset interval into regions where one or two streams are perceived, as well as an ambiguous region in which perception is dependent on attention (diagram available in [7]). While the van Noorden diagram is typically used to map stream segregation with respect to pitch differences, Barass et al. have broadened the palette of such diagrams to other audio properties including brightness of a noise grain, and then to amplitude and inter-level difference panning of the noise grain [5]. van Noorden diagrams can be useful in guiding multi-stream sonification algorithms away from regions of perceptual ambiguity.

Spatialization can be used to minimize interference and masking effects in audio. The use of spatialization to regulate stream segregation has a number of specific advantages. Firstly, it is an independent parameter in the sense that the spatialization of a particular stream can be modified without risk of the change interfering with other parameters (conversely, if e.g. pitch and timbre are modified independently, interaction between these two dimensions may cause ambiguity in the resulting stream) [8]. Secondly, spatialization provides the user with a physical map for the sound sources which can aid in the sonic learning process.

The capacity for spatialized audio to aid in the perception of segregated streams has been evaluated only a handful of times in prior work. A study that is most closely related to the current discussion compared user identification of two pitch-segregated signals that are spatially segregated with two pitch-segregated signals emanating from a single source. Researchers found improved signal classification accuracy when audio streams are spatially segregated, among other findings described in [9].

3.3. Data Sonification Platforms

Up until now our discussion has focused predominantly on theory and prior work in the study of auditory perception. Rotator, in more practical terms, is intended as a sonification platform.

A number of generalized data sonification platforms have been developed in the past. Four examples are the Sonification Sandbox (a Java application developed by the Georgia Institute of Technology) [10], SonArt (a platform developed at Stanford University) [11], Personify (a scientific data sonification platform built at CSIRO) [12], and MUSE (A musically-driven data sonification platform created by UC Santa Cruz) [13].

Each tool makes use of visual information to aid in the user experience. For example, Personify asks users to customize a visual representation of a data-to-audio mapping space where axes correspond to musical properties, and the Sonification Sandbox allows users to view line plots of data streams as they are being sonified. However, unlike the sonification platforms cited above, the Rotator platform is specifically aimed at diversifying the way that users *distribute* data across their senses, which we have yet to see as the focus area of a sonification tool.

4. SYSTEM ARCHITECTURE

4.1. Architectural Overview

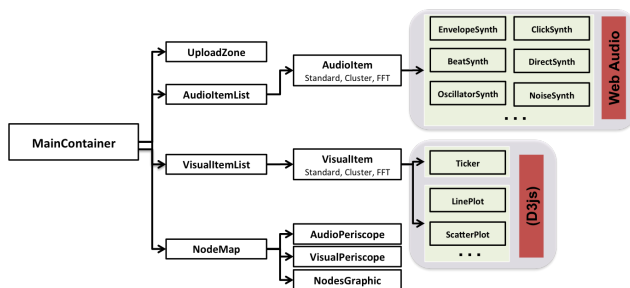


Figure 3: Simplified React view component hierarchy for the Rotator application. Note that additional communication loops exist that are not pictured

Rotator is a client-side application built in Javascript and architected in React and Flux. The most heavily used low-level Javascript libraries are D3 (for visualization) and Web Audio (for sonification). The decision to build Rotator as a web-based application was reached for a few principle reasons. Firstly, real-time, web-based audio technology has improved tremendously over the last decade and is rapidly becoming a standard. Secondly, previous experience integrating together common audio processing tools like Pure Data, Max/MSP, and Ableton Live suggest that the

installation process for a new user can be cumbersome; a web-based application requires only a browser. Finally, extending a niche visualization tool such as ROOT¹ was entertained as a possibility but would target a narrower user base than a web-based tool. Web Audio does have some drawbacks, however. Most notably, it is still under active development and therefore lacks certain basic audio streaming features that are commonplace in desktop audio processing software, e.g. adjustable sampling rate. Secondly, Web Audio is yet to be fully ported to React and must therefore be integrated by the developer. (Note that [15] is a first pass at Web Audio integration into React, but is yet to support audio spatialization).

Figure 3 shows a simplified flowchart of the React view components designed for the Rotator application. There are three principle data flow chains: one controlling sound synthesis, one controlling visualizations, and one controlling the node map used to navigate within a dataset. A combination of callback functions and Flux data store updates enables communication between sibling components. Further, a custom audio scheduler was built for audio synchronization across Rotator components.

4.2. Synthesizer Architecture

In the current software architecture, each data stream’s associated synthesizer component is mounted and activated when the user wishes to sonify the stream. Bearing in mind principles in psychoacoustic theory described in Section 3, six synthesizer types have been built as default options: **noise**, **envelope**, **clicks**, **oscillator**, and **beats**, and **direct**. Each synth’s parameters are controlled via user-defined values supplied in the lefthand control panel previously pictured in Figure 2. In addition, it is possible for a user to integrate an entirely new audio synthesizer component as long as it accepts a set of React props required by all Rotator synthesizers.

The white noise synth populates an audio buffer of size 100,000 with random values and then channels the buffer through a low-pass filter with a cutoff frequency driven by the data. The envelope synthesizer defines an envelope’s attack, decay, sustain, and release times, ramping up an oscillator to a frequency driven by the incoming data. The click synth is another iteration of the envelope synth but with parameters shorter in duration, and applied to a buffer filled with white noise. The oscillator synth uses incoming data to modulate the pitch of the oscillator. The beat synth creates a beat effect between two oscillators, using data to modulate the small frequency gap between each oscillator. The further that the incoming data is from a preset threshold value, the higher the frequency of the beats. Basic wave shaping distinguishes data values above and below the threshold; values below the threshold trigger two sinusoidal oscillators, and values above the threshold trigger two triangle oscillators. Finally the direct synth creates a direct audification of the data by populating an audio buffer that is played back at a rate determined by the user-defined tempo. Audio samples can be accessed at <http://resenv.media.mit.edu/rotator>.

5. APPLICATION AREAS

Three application areas for Rotator were implemented. Two scenarios (quantum algorithm interpretation, temperature monitoring in an experiment) will be briefly described. The most promising scenario (biosensor data analysis) is described more thoroughly and is incorporated into a user study described in Section 6. Figure 4 shows the node maps used while testing each scenario.

¹ROOT is a C++ visualization framework used in particle physics [14].

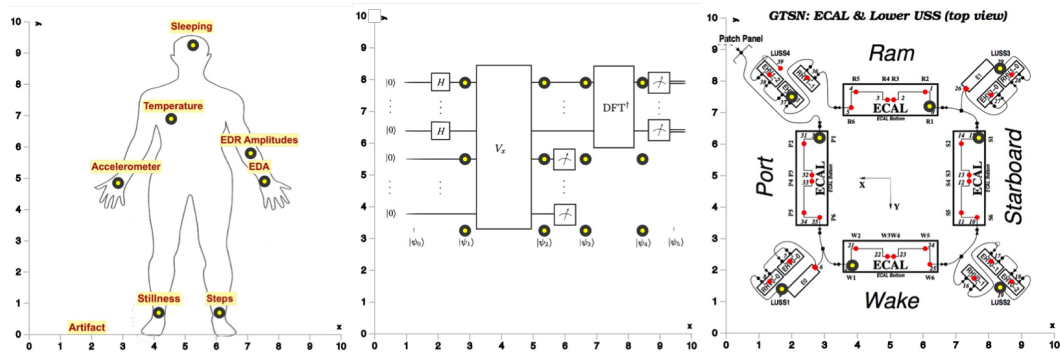


Figure 4: Auditory and visual windows can be slid and resized across appropriate node maps, as in Figure 1. Maps for the three application areas considered are pictured: biosensor data (left) quantum algorithm circuit (center) temperature monitoring in an experiment (right)

5.1. Quantum Algorithm Interpretation

Audio may be a very suitable data display mode for quantum systems. This presumption is due in large part to a phenomenon known as quantum superposition, which describes the capacity for a quantum bit to exist in two states at once (with some probability of a measurement yielding either state). Similarly, as described in section 3.2, audio can be layered in a way that triggers the simultaneous perception of multiple streams, and is thus a natural display mode for data encoded in a quantum system. Rotator was used to sonify the quantum states that arise in a famous quantum factoring algorithm called Shor's algorithm [16]. We were specifically intrigued by the possibility for Rotator to immerse a user in different stages of the algorithm simultaneously. For instance, using Rotator, it is possible to hear an algorithm state prior to a quantum operation in one's left ear and an algorithm state following quantum operation in one's right ear, while visually examining plots of the remaining stages of the algorithm. Furthermore, a quantum algorithm's circuit diagram provides a natural node map within Rotator that can guide user interaction.

Shor's algorithm is an interesting choice for sonification. On the one hand, it is a quintessential quantum algorithm with a periodic structure that can be elucidated sonically. Oscillator synths were used to convey periodic states and click synths were used to convey the results of quantum measurements on the states. On the other hand, the high quantum bit numbers per register that are required even for factoring small N result in a very large superposition of states to work with. The states have equal probability amplitude at many stages of the algorithm (as compared to e.g. the quantum harmonic oscillator where a time-dependent term modulates the probability amplitudes as the system evolves), and the important periodic structure in the algorithm does not lie in the probability amplitudes of the states, but instead in the series of states themselves. To make this periodicity apparent in audio, we must iterate through the states in superposition rather than play them simultaneously, which breaks from our initial motivation to sonify quantum systems in the first place! Be that as it may, the current sonification approach is more generalizable across classical algorithms. Furthermore, for very low N , it may be possible to sonify the states of all bits in a register simultaneously rather than iterate through each bit state, which is worthy of further consideration. However, to attain the full benefit of sonic display of quantum data, it is recommended that future approaches select quantum systems with small numbers of states and with a time dependent parameter

that modulates the probability amplitudes of these states.

5.2. AMS Temperature Data

The alpha-magnetic spectrometer (AMS) is a physics experiment onboard the International Space Station (ISS) responsible for detecting particles that may contribute to our understanding of dark matter, anti-matter, and cosmic rays [17]. Currently, hundreds of sensor node readings are presented as tables in the AMS control room and are color coded in red, green, and yellow on the basis of expected vs. anomalous behavior.

A small, offline dataset consisting of 8 temperature nodes located throughout the detector was provided for consideration of alternative data display modes, and an initial sonification scheme was developed. Rather than aim for maximal stream segregation, an array of temperature nodes lends itself to representation as a soundscape that blends together, creating a feeling in the listener of being immersed in a heat map. Therefore, a simple mapping was created using eight spatialized oscillator synths, with temperature controlling the pitch of each synth. As users grow accustomed to this ambient hum, they will be able to identify warmer and cooler regions of the detector while growing sensitive overtime to perceived anomalous behavior, which can then be validated visually. Rotator's node clustering feature was of particular use in this application area, since each region of the detector behaves similarly.

The AMS temperature dataset is particularly appealing due to its behavior at multiple timescales: high-frequency oscillations are caused by 92-minute orbitals, and low frequency oscillations are caused by gradual drift in ISS orbit. In order to perceive both scales at once using visual tools, it is necessary to either zoom in and out, else look back and forth between two plots at different time scales. A tool like Rotator can be used to provide the user a sonification of the high frequency oscillations of a temperature stream, while an accompanying visualization simultaneously provides the user with larger timescale, low frequency behavior. This approach is most readily accomplished in Rotator by increasing the modulation depth of a stream's synthesizer in order to amplify smaller-scale changes in the sound. Meanwhile, a lowpass-filtered signal can be fed in by the user for visualization. In this way, the user is able to simultaneously track variations of a temperature node at two different timescales, while only monitoring a single auditory and single visual track.

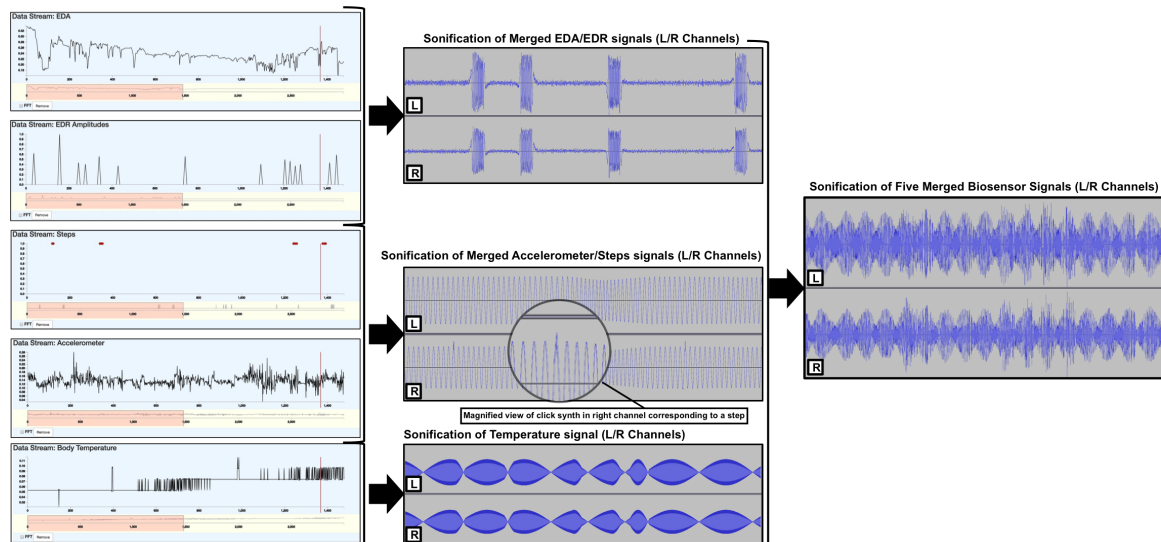


Figure 5: Sample sonification of five biosensor data parameters. The left-most five plots contain excerpts from each data stream as visualized in Rotator (note that the pink bar at the base of each plot is resizable and slidable to adjust data range displayed). Three audio streams depict audio from EDA and extracted EDRs (top), accelerometer and extracted steps (middle) and temperature (bottom). The rightmost plot shows the resulting audio stream when all five parameters are sonified at once.

5.3. Biosensor System Data

Data from one participant of a largescale user study on wellbeing was provided as a testbed for the Rotator tool (see [18] for study information). The participant of this study was equipped with the Affectiva Q wrist-worn physiological sensor that records acceleration, skin temperature, and Electrodermal Activity (EDA).

The principle task of the researchers who own this dataset is to develop models that predict a person’s stress, happiness, and ultimately, depression. Because physical activity (steps), sedentary activity (stillness), and movement speed are all relevant to depression, the researcher must correctly extract these features from the data. Furthermore, a peak with a particular waveshape in the EDA signal represents an electrodermal response (EDR), which occurs due to increased sympathetic nervous system activity and can be indicative of increased emotion and stress; therefore EDRs must be correctly detected as well. However, EDRs that occur during movement or when the person’s body temperature is high are less likely to relate to stress. Thus an understanding of the interrelations between multiple features and signals plays a central role in deducing how to best analyze the data. Rotator can be used to test flexible audio-visual representation as an alternative approach for interpreting high-dimensional data.

Rotator was originally designed for applications in which data derived from a network of sensors is analyzed. However, the dataset under discussion is drawn exclusively from a single wrist-worn sensor. Therefore, a sensible geometric layout was imposed in which each data node is spread across a human body contour line (Figure 4, left). A node map associated with a physical body enables spatialization of streams with respect to a reference map, and is thought to expedite the training process for new users.

The Biosensor dataset lends itself well to sonification in that many data types are present: transients (steps, EDR), continuous rapidly varying signals (accelerometer, EDA) continuous, slow-changing signals (body temperature), and binary signals (artifact

detection). A sonification scheme designed for five features is summarized in Table 1 and an example of sound waves juxtaposed with their corresponding raw data streams is shown in Figure 5. Clicks and pulses are used for the transient data, which are particularly readily perceived when spatialized with respect to the listener. The EDA signal is sonified as white noise with a cutoff frequency modulated by the data. The user may intuitively draw a connection between the rushing-water sound of white noise and the activity level of the skin. Furthermore, it has been demonstrated in prior literature that transients and noise are most readily distinguished by the listener. Recalling discussions from Section 3, this phenomenon is largely due to the fact that transients are most likely to have distinctive onset times. Next, the accelerometer data is sonified as a high-frequency-band oscillator with modulated frequency. The rapidly changing nature of the raw accelerometer data lends itself well to this simple oscillator audification. Body temperature was sonified using a low-frequency-band beat effect synth with beat frequency corresponding to deviation from average, and wave shaping used to distinguish between positive and negative deviations. Thus, the user can determine how far body temperature has drifted, and in what direction. The frequency of the temperature synth is segregated from that of the accelerometer synth by at least 6 semitones, bearing in mind the van Noorden diagram temporal cohesion boundary for pitch segregation (see [7]). Taken together, the user is able to derive a feeling for the stress and activity levels of the person under study by growing acquainted with the interplay between variables as well as validating hypotheses by shifting data between their visual and auditory senses.

6. EVALUATION

6.1. Overview

A small-scale user study was conducted in order to gauge the influence of flexible audio-visual data display on a user’s ability to

Accelerometer	Oscillator with modulated frequency
EDA	Filtered white noise with modulated cutoff
Temperature	Beat effect + wave shaping
Steps	Clicks generated using envelope function
EDR	Oscillator-driven pulse

Table 1: Biosensor Sonification Scheme Summary

draw conclusions about the structure of a dataset. In particular, the study is designed to measure a user’s perceived cognitive load under different perceptual conditions, as well as measure the speed and tentative accuracy with which the user completes a task. The Biosensor dataset was chosen for use in this study because it included the most varied and developed sonification mappings of the three application areas considered in Section 5.

6.2. Methodology

The biosensor dataset consists of five parameters listed in Table 1 and drawn from one day of stressful activity and one day of calm activity. Participants were asked to classify approximately six-minute-long excerpts² from the time series data set on a 5-point ranking of stressed/relaxed and active/still. The users were asked to perform this task for between five and ten data samples under four perceptual conditions, which all users experienced in the following order: (1) all streams presented as audio (‘A’), (2) all streams presented as plots (‘V’) (3) all streams presented as audio and up to one stream at a time simultaneously presented visually (‘AVS’) (4) A subset of streams presented as audio and up to one stream at a time presented visually; the visual stream *cannot* also be sonified (‘AVD’). Conditions (3) and (4) distinguish between a scenario in which the user is observing a plot that they simultaneously hear and a scenario in which the user is seeing plots that they do not simultaneously hear. As long as the conditions of each trial were met, the users were free to move around the audio and visual boxes (see Figure 1 as a reminder) as well as adjust the synthesizer mapping parameters available in the lefthand panel. After rating each sample on the basis of stress level and activity level, the user presses a button to load the next sample. Browser localStorage was used to collect responses.

Each of five participants spent approximately two hours using the Rotator platform. Participants did not have any hearing impairments. Audio was spatialized using Web Audio’s head related transfer function (HRTF) panning model and presented to the user via traditional headphones. The first ~45 minutes served as a training session. The training session consisted of the following steps:

- Verbal overview of per-stream data-to-audio mapping
- Review sample audio clips and visualizations for each stream
- Review Table 2, a qualitative guide for assessing stress and activity levels based on available data
- Review sample audio clips and visualizations for extreme states: high-stress, low-stress, high-activity, and low-activity
- Review samples of all five audio streams playing at once

If a data sample is deemed to meet the second criterion in Table 2 (‘emotional’), then the participant was instructed to rate the

²Note that the user could set arbitrary rates for scanning the data

Temperature	EDA/EDR	Accel/Steps	State
Low	High	Low	Stressed
High	High	Low	Emotional
High	Either	High	Active

Table 2: Excerpt from an information sheet provided to user study participants for characterizing data samples

user as having some moderately increased stress level. It is important to note that there are many biological caveats to the heuristics provided in Table 2. For the purpose of a perceptual study, it is only important for the provided identification instructions to be consistent across all user study participants.

After completing trials under each of the four perceptual conditions, users were asked to fill out a NASA task load index (NASA-TLX) survey, a widely used assessment tool for ranking perceived cognitive workload to complete a task [19]. Finally, at the end of the study, users filled out a final survey regarding their subjective experience performing the task under each condition.

6.3. Results

The rounded and weighted NASA TLX scores for each participant after completing trials in each sensory mode are provided in Table 3. We make a few key observations from the results in Table 3: firstly, the participants’ self-reported prior audio experience correlates in all 5 cases with their perceived task load ranking for the audio-only trial. Note that while in all cases, NASA-TLX cognitive load measurement for the auditory-only scenario were highest, 4 out of 5 users also stated that of all the conditions, their performance in the all-audio condition seemed to improve the most (this result, a survey question, is not pictured). This result reaffirms the importance of a training period in auditory display comprehension. Furthermore, we note that users remained most comfortable and efficient in the visual-only mode. However, there were only increases in perceived task ease in the auditory modes, which also suggests that additional training may affect the ‘AVS’ vs. ‘V’ comparison.

Secondly, and most critically, we observe in Table 3 that for 4 out of 5 participants, the ‘AVS’ task was ranked as requiring nearly equivalent cognitive load as the all-visual task (the ‘AVS’ task, as a reminder to the reader, is the task in which all audio was sonified but one stream at a time could also be visualized, functioning as a ‘peeking’ mode). For the one participant for whom this was not the case, the ‘AVS’ task was still ranked as the easiest to complete among the three modes involving audio (NASA TLX of 69 compared to NASA TLX 77 for ‘AVD’ and NASA TLX of 83 for ‘A’). Furthermore, this subject verbalized extreme lack of familiarity working with audio. The post-experiment survey provides further validation for the ‘AVS’ mode: 4 out of 5 users included the ‘AVS’ state among trials in which they felt that their performance improve over time (as compared to 3 out of 5 users indicating improvement in the ‘AVD’ and ‘A’ modes and 1 out of 5 users indicating improvement in the ‘V’ mode).

There are a number of confounding factors to consider. Firstly, participants are only exposed to one possible visualization methodology, so it is possible that a different visualization approach would impact results. Secondly, since the ‘A’ task was performed prior to the ‘AVS’ and ‘AVD’ tasks for all participants, it is possible that participants grew more accustomed to the sonification during the

Participant	TLX ('A')	TLX ('V')	TLX ('AVS')	TLX ('AVD')	Audio Experience?
1	61	51	43	n/a	A lot
2	69	42	46	52	A lot
3	73	33	33	48	None
4	58	51	50	54	Some
5	83	38	69	77	None

Table 3: Summary of NASA TLX weighted scores for each of 5 user study participants under four different perceptual modes, which self-described audio experience marked for each participant. TLX scores are out of 100, with higher TLX scores indicate greater perceived cognitive workload, and lower scores indicate less perceived cognitive workload. See body of text for descriptions of perceptual modes

	1 (easiest)	2	3	4	5 (hardest)
clicks	5	0	0	0	0
white noise	0	3	0	2	0
envelope	0	1	3	0	1
oscillator	0	1	2	2	0
beat effect	0	0	0	1	4

Table 4: Ranking of the five synthesizers on the basis of how readily identifiable they are when played in tandem. 1 is easiest to identify, 5 is hardest to identify

latter tasks. However, based on textual user reports, the visual peeking feature in the 'AVS' mode was particularly useful as a means of validating any auditory cues. One participant writes 'I definitely felt the most confident in the 'AVS' scenario, because I could get a quick sense from the audio and then allay any questions/concerns with a few targeted visual queries.' Another writes: 'The visual information was especially helpful in verifying what I was hearing.' Users in the 'AVS' mode opted for varying number of visual validations using the sliding visual window based on the audio sample under examination. A quantitative study comparing how audio properties influence user interaction with the sliding windows could be worthwhile future work to undertake.

6.4. Evaluation of Synthesizers

Users were asked to rank each of the 5 synthesizer types on the basis of how easily identifiable they are when played in synchrony. Rankings are shown in Table 4. Despite discrepancies in prior audio experience and perceived task workload, there was great consistency among the rankings. All users rated the steps as easiest to identify (sonified using the clicks synth) and 3/5 users rated the EDA second easiest to identify (sonified using the noise synth). Noise and transients are known to be particularly readily distinguished from one another. On the other hand, temperature data, sonified using the beat synthesizer, was least readily identifiable by participants. A plausible explanation involves the additional steps required to learn the data encoding scheme used for beat synthesizer e.g. remembering what high and low beat frequencies indicate, as well as identifying the two wave shapes corresponding to above-average and below-average. Perhaps with additional training this stream would become more readily identifiable. As a qualitative note, users appeared to only occasionally adjust the frequency and gain controls for each synth and did so mainly as a mechanism for studying the sonifications, according to post-experiment surveys. Otherwise users generally preferred to keep the controls fixed and rely on the sliding windows for adjustments.

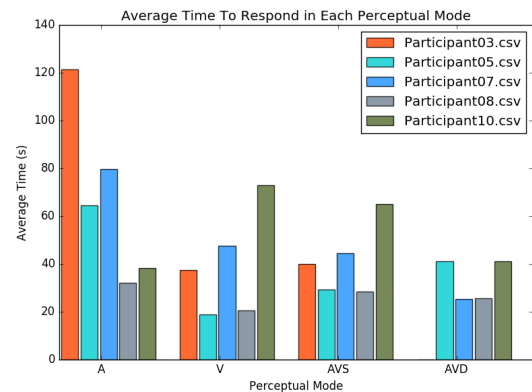


Figure 6: Ranking of average time spent in each mode for each participant. Participant 03 did not complete the 'AVD' trials; participants 03 and 07 self-reported as having more audio experience

6.5. Evaluation of Task Completion Time

Many users requested that future studies include additional training time beyond the allotted 45 minutes, and in particular would appreciate immediate performance feedback during the training. However, there were no discernible trends in task completion time over multiple trials. Furthermore, Figure 6 shows that for most participants, the time required to complete the dual audio-visual tasks was comparable to completion time for the purely visual task. Figure 6 also shows that for participants with more experience working in audio (participants 03 and 07), auditory tasks took longer to complete despite taking less subjective cognitive effort according to the NASA-TLX. This may be due to an increased capacity for discerning subtlety in audio.

6.6. Evaluation of Task Accuracy

It is difficult to verify the accuracy with which users completed each task since there is no source of ground truth for the stress or activity level represented by the data. However, we studied various coarse accuracy measurements. For example, the number of detected steps serves as the closest coarse 1-dimensional measure for activity level. However, recalling Table 2, if users ranked purely according to step count, they would be failing to take into consideration the additional biosensor parameters. Therefore, assessments of this sort should be treated as very approximate. Figure 7 shows how users ranked activity level based on step count across all available trials for each sensory mode. The audio-only mode

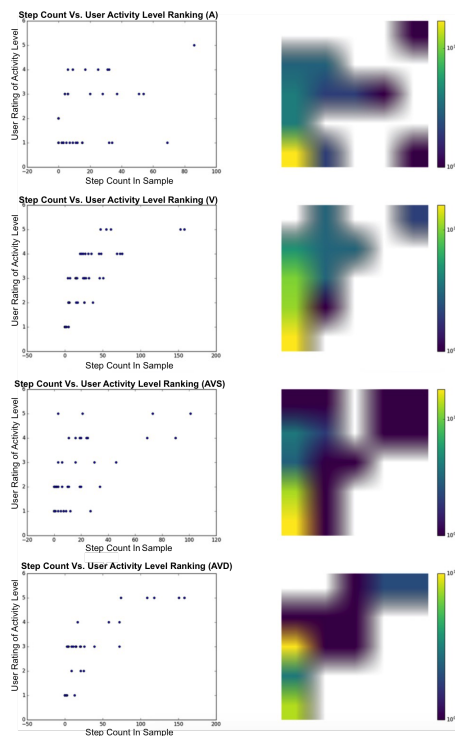


Figure 7: Scatter plot and heat map showing step count (x-axis) plotted with respect to user ranking of activity level (y-axis) for modes ‘A’ (top) ‘V’ (second) ‘AVS’ (third) and ‘AVD’ (bottom)

shows the most spread in ratings, suggesting the highest degree of perceptual ambiguity. The remaining 3 modes show very low false-positives in user perception, demonstrating the capacity for a dual audio-visual mode to improve perception.

Further, a cursory application of linear discriminant analysis across the dataset suggests that while high step count correlates most strongly with high activity rating in each of the four modes, high temperature count correlates more strongly with high activity rating only in the three modes containing auditory feedback, our first hint at improved performance in auditory modes.

7. SUMMARY AND FINAL THOUGHTS

Rotator has been used to enable a study of flexible audio-visual presentation modes for data display. One promising conclusion is the noteworthy jump in perceived task ease, task completing time, and tentatively in performance accuracy for the ‘AVS’ presentation mode as compared to the audio-only presentation mode. It is quite likely that optimal presentation modes are data and task dependent. Thus, platforms that provide flexibility of data stream display mode may be most broadly applicable. As a final thought, perhaps geometrically high-dimensional data can be better represented in an n-dimensional audio-visual space, with high-dimensional rotations used to adjust both sonic and visual projections.

8. REFERENCES

- [1] W. Barfield, C. Hendrix, O. Bjorneseth, K. A. Kaczmarek, and W. Lotens, “Comparison of human sensory capabilities with technical specifications of virtual environment equipment,” *Presence: Teleoperators & Virtual Environments*, vol. 4, no. 4, pp. 329–356, 1995.
- [2] J. J. Todd, D. Fougny, and R. Marois, “Visual short-term memory load suppresses temporo-parietal junction activity and induces inattentive blindness,” *Psychological Science*, vol. 16, no. 12, pp. 965–972, 2005.
- [3] M. Czerwinski, N. Lightfoot, and R. M. Shiffrin, “Automatization and training in visual search,” *The American journal of psychology*, pp. 271–315, 1992.
- [4] P. Cariani and C. Micheyl, “Toward a theory of information processing in auditory cortex,” in *The Human Auditory Cortex*. Springer, 2012, pp. 351–390.
- [5] S. Barrass and V. Best, “Stream-based sonification diagrams,” in *Proceedings of the 14th International Conference on Auditory Display*, 2008.
- [6] A. S. Bregman, *Auditory scene analysis: The perceptual organization of sound*. MIT press, 1994.
- [7] F. Almonte, V. K. Jirsa, E. W. Large, and B. Tuller, “Integration and segregation in auditory streaming,” *Physica D: Nonlinear Phenomena*, vol. 212, no. 1, pp. 137–159, 2005.
- [8] E. J. e. a. Allen, “Symmetric interactions and interference between pitch and timbre,” *The Journal of the Acoustical Society of America*, vol. 135, no. 3, pp. 1371–1379, 2014.
- [9] H. J. Song and K. Beilharz, “Concurrent auditory stream discrimination in auditory graphing,” *Journal of Computers*, vol. 3, pp. 79–87, 2007.
- [10] B. N. Walker and J. T. Cothran, “Sonification sandbox: A graphical toolkit for auditory graphs,” in *Proceedings of the 2003 International Conference on Auditory Display*, 2003.
- [11] O. Ben-Tal, J. Berger, B. Cook, M. Daniels, and G. Scavone, “Sonart: The sonification application research toolbox,” in *Proceedings of the 2002 International Conference on Auditory Display*, 2002.
- [12] S. Barrass, “Personify: a toolkit for perceptually meaningful sonification,” in *ACMA*. Citeseer, 1995.
- [13] I. A. Popa, J. E. Boyd, and D. Eagle, “Muse: a music-making sandbox environment for real-time collaborative play,” *perspective*, vol. 3, p. 4, 2015.
- [14] “Root data analysis framework,” retrieved August 19th, 2016 from <https://root.cern.ch/>.
- [15] G. Haussmann, “react-webaudio,” retrieved August 19th, 2016 from <https://github.com/Izzimach/react-webaudio>.
- [16] C. Lavor, L. Manssur, and R. Portugal, “Shor’s algorithm for factoring large integers,” *quant-ph/0303175*, 2003.
- [17] S. Ting, “Alpha magnetic spectrometer,” in *39th COSPAR Scientific Assembly*, vol. 39, 2012, p. 1975.
- [18] “Snapshot study,” granted data sample by Natasha Jaques and Sara Taylor. Study summary is available at <http://snapshot.media.mit.edu/>.
- [19] S. G. Hart and L. E. Staveland, “Development of nasa-tlx (task load index): Results of empirical and theoretical research,” *Advances in psychology*, vol. 52, pp. 139–183, 1988.